Estimates of Gene Flow Among Populations, Geographic Races, and Species in the *Ipomopsis aggregata* Complex

Paul G. Wolf[1] and Pamela S. Soltis

Department of Botany, Washington State University, Pullman, Washington 99164-4238

Manuscript received May 9, 1991
Accepted for publication November 9, 1991

ABSTRACT

Interpopulational gene flow within a species can reduce population differentiation due to genetic drift, whereas genetic exchange among taxa can impede speciation. We used allozyme data to estimate gene flow within and among geographic races and species of perennial herbs in the *Ipomopsis aggregata* complex (Polemoniaceae). Estimates of interpopulational gene flow within taxa from two methods (*F* statistics and private alleles) were correlated with one another. Gene flow among populations within each geographic race (subspecies) of *I. aggregata* was relatively high ($Nm > \sim 1.0$). Gene flow was also high among populations of *I. arizonica* and among four northern populations of *I. tenuituba*. However, gene flow was low ($Nm < 1.0$) for *I. tenuituba* when a population representing subsp. *macrosiphon* was included. This is consistent with previous findings that subsp. *macrosiphon* has had an independent origin and is reproductively, as well as geographically, isolated. A recently developed model, based on hierarchical *F* statistics, was employed to estimate genetic exchange among taxa. Gene flow estimates were generally high among races of *I. aggregata* ($dNm_{race} > 1.0$) but were low among subspecies of *I. tenuituba* ($dNm_{race} < 1.0$). Consistent with morphological evidence, estimates of interspecific gene flow were moderate between *I. aggregata* and *I. tenuituba*, which hybridize in several areas. However, contrary to morphological evidence, we estimated relatively high levels of interspecific gene flow involving *I. arizonica*. Our results suggest that *I. arizonica* has hybridized with other species without the transfer of morphological traits. In the *I. aggregata* complex, gene flow appears to be an important evolutionary force shaping geographic variation for allozymes within species, but is insufficient to prevent morphological divergence among taxa.

GENETIC exchange among natural populations has evolutionary implications both above and below the species level. According to the classical view, gene flow among populations of a species is an essential factor in maintaining species integrity (MAYR 1970), although the relative importance of this reproductive continuity has been questioned (EHRLICH and RAVEN 1969; PATERSON 1985; TEMPLETON 1989). Correspondingly, the cessation or near-cessation of gene flow is implicit in most models of speciation, in which subsequent genetic drift, selection, or both, result in divergence (*e.g.*, STEBBINS 1950; MAYR 1963; GRANT 1981).

The importance of gene flow to the speciation process has been approached from the context of interspecific hybridization in plants (LEVIN 1975) and animals (BUTLIN and HEWITT 1985; HEWITT 1989). However, hybridization is a postspeciation event, and although it may cause the breakdown of a species boundary it may not help to explain the origins of species differences. Here we focus on the amount of gene flow that can occur without impeding speciation. If two groups of populations are exchanging genes, the main force acting to prevent speciation is recombination, which dismantles the association between genes controlling reproductive isolation and those conferring adaptations to different niches (FELSENSTEIN 1981). In laboratory stocks of *Drosophila melanogaster*, genetic divergence for simple traits can occur despite maximum (50%) gene flow between the diverging populations (THODAY and BOAM 1959; STREAMS and PIMENTAL 1961). However, it is not known how much gene flow can occur without completely hindering the speciation process in nature.

Our approach to the above question involves an analysis of gene flow statistics in an actively speciating plant group that includes a continuum from geographic races to presumably distinct "biological" species. Despite extensive discussions of the theory, few empirical studies have attempted to examine the role of gene flow during speciation (*e.g.*, HARRISON 1979; MALLET *et al.* 1990). Recently, PORTER (1990) used indirect estimates of gene flow to examine the status of genetic isolation between hybridizing butterfly species. Thus, gene flow statistics can also provide useful information for systematists interested in species boundaries.

If individuals of a species are clustered into distinct populations, gene flow between populations can be described as the product of the effective population

[1] Present address: Department of Biology, Utah State University, Logan, Utah 84322.

size (N) and the proportion of migrants per generation (m). If populations are of similar size, then Nm describes the average number of individuals per generation migrating between populations. The significance of Nm was first considered for an island model by WRIGHT (1931), who determined theoretically that if more than one individual migrates between populations every other generation ($Nm > 0.5$), different alleles at a locus in the two populations would not become fixed due to genetic drift. Thus, gene flow is the major determinant of population structure when $Nm > 0.5$, and significant genetic differentiation can result from genetic drift when $Nm < 0.5$.

Interpopulational gene flow can be estimated directly from observations of dispersal or movement of marker genes. Such an estimate gives a measure at one, or a few, points in time. Alternatively, allele frequencies can be used to infer average levels of gene flow indirectly (SLATKIN 1987). Many studies have made such indirect estimates, and data compiled from studies of plants (HAMRICK 1987) and animals (SLATKIN 1985) indicate considerable variation among species, with Nm values ranging from 0.26 in *Phlox cuspidata* (based on LEVIN 1975) to 42.0 in the marine bivalve, *Mytilus edulis* (SLATKIN 1985).

Herein, we estimated interpopulational gene flow within and among taxa in the *Ipomopsis aggregata* complex (Polemoniaceae). This group consists of outcrossing, perennial herbs found throughout western North America. GRANT (1981) first presented the *I. aggregata* complex as an example of incipient speciation. In the most recent taxonomic revision, GRANT and WILKEN (1986) treat the group as a three-species complex with *I. aggregata* as the most widespread species for which seven geographic races (treated as subspecies) are recognized. *Ipomopsis tenuituba*, a subalpine species with three geographic races, hybridizes with *I. aggregata* in some areas, but the two species grow sympatrically without hybridizing in other regions. The third species, *Ipomopsis arizonica*, does not hybridize with the other taxa in nature according to morphological evidence (GRANT and WILKEN 1986). All taxa in the complex can be artificially crossed, but F_1 hybrids involving *I. arizonica* and either *I. aggregata* or *I. tenuituba* have reduced fertility (D. WILKEN, unpublished data). The primary pollinators in the group are hawkmoths and migratory hummingbirds, making the potential for gene flow via pollen considerable. However, seeds of Ipomopsis have no obvious mechanisms for dispersal and are not moved long distances (WASER and PRICE 1983).

We used allozymic data to make indirect estimates (SLATKIN 1987) of gene flow among populations within geographic races and species, and among races and species. All estimates are based on historical patterns of gene flow and reflect the amount of gene flow in the recent past that is assumed to have resulted in the present-day distribution of genetic variation.

TABLE 1

Numbers of populations sampled and distribution of taxa

Species	Subspecies (=race)	No. of populations sampled	Distribution[a]
I. aggregata	*aggregata*	6	ID; MT; UT(n); CO(w)
	formosissima (n)	5	WA; OR; CA(n)
	formosissima (s)	7	NM; AZ; UT(s)
	attenuata	6	CO(n); WY(s)
	bridgesii	4	CA
	collina	6	CO; NM(n)
	candida	7	CO; NM(n)
	weberi	4	CO(n); ID(w)
I. tenuituba	*tenuituba*	2	CA; OR; ID(s); UT; CO
	latiloba	2	UT(s); AZ(n)
	macrosiphon	1	AZ(s)
I. arizonica		6	UT(s); AZ(n)

[a] States are abbreviated; letters in parentheses indicate specific region of state: n = north, etc.

MATERIALS AND METHODS

Fifty-six populations were sampled across the range of the *I. aggregata* complex. Numbers of populations sampled and distributions of taxa are listed in Table 1. Genotype data were obtained at 23 allozyme loci (18 of which were polymorphic) from 30 plants per population. Locality information, sampling procedures and electrophoretic protocols are reported elsewhere (WOLF, SOLTIS and SOLTIS 1991), and allele or genotype frequencies can be obtained from P. G. WOLF by request.

Two approaches were used to estimate gene flow; the first is based on the relationship between Nm and the degree of population substructuring as determined by WRIGHT (1951):

$$Nm = \{(1/F_{ST}) - 1\}/4 \quad (1)$$

where F_{ST} is defined as the standardized genetic variance among (sub)populations in the total sample (WRIGHT 1965). This relationship between F_{ST} and Nm assumes that populations are arranged according to an island model. However, SLATKIN and BARTON (1989) showed that the relationship is relatively robust to deviations from the island model. An advantage of estimating Nm from F statistics is that gene flow can be estimated at different levels in a hierarchy. WRIGHT (1978) developed an analysis of genetic differentiation whereby the degree of genetic subdivision could be partitioned at several levels. When applied to the three-tier hierarchy in the present case:

$$1 - F_{\text{pop.total}} = (1 - F_{\text{pop.race}})(1 - F_{\text{race.sp}})(1 - F_{\text{sp.total}}). \quad (2)$$

Gene flow among groups in a two-tier hierarchy can be estimated from hierarchical F statistics following SLATKIN and VOELM (1991). This can be extrapolated to a three-tier hierarchy, with k species, each of n races, each of d populations:

$$F_{\text{pop.race}} = \frac{1}{1 + 4Nm_{\text{pop}} \dfrac{d^2}{(d-1)^2}} \quad (3)$$

$$F_{\text{race.sp}} = \frac{1}{1 + 4Nm_{\text{race}}\, d\, \dfrac{n^2}{(n-1)^2}} \quad (4)$$

$$F_{\text{sp.total}} = \frac{1}{1 + 4Nm_{\text{sp}}\, dn\, \dfrac{k^2}{(k-1)^2}} \quad (5)$$

where m_{pop} is the migration rate between populations of the same race, m_{race} is the migration rate among populations in different races of the same species, and m_{sp} is the migration rate among populations from different species.

SLATKIN and BARTON (1989) discuss the various algorithms for estimating F statistics (e.g., NEI 1973; WEIR and COCKERHAM, 1984) and conclude that the method used will not strongly affect the estimate of Nm. However, PORTER (1990) found that estimates of gene flow among taxa were seriously overestimated using the method of NEI (1973), and values calculated following WEIR and COCKERHAM (1984) were more realistic, based on known sibling species relationships. In this study we have also estimated F statistics using both methods. We used genotype frequency data to calculate G_{ST} [the (NEI 1973) estimator of F_{ST}], employing the computer program BIOSYS-1 (SWOFFORD and SELANDER 1981) to generate hierarchical F statistics following the WRIGHT78 step. A modification of BIOSYS-1 (W. BLACK, personal communication) was used to calculate θ [the WEIR and COCKERHAM (1984) estimator of F_{ST}]. Values of θ were based on means from jackknifing over loci. Hierarchical F statistics were then used to make estimates of gene flow at different levels in the hierarchy. For interspecific F statistics and gene flow, *I. aggregata* was analyzed in two ways: (1) we used only the southern group of seven populations of *I. aggregata* subsp. *formosissima* to represent *I. aggregata* (*aggregata*$_7$). This provided sample sizes of similar magnitude among species and used the populations of *I. aggregata* most sympatric with *I. arizonica* and *I. tenuituba*. (2) Calculations were also performed using all 45 populations of *I. aggregata* samples (*aggregata*$_{45}$). Gene flow among races was estimated for the entire sample and also separately for *I. aggregata* and *I. tenuituba*. Likewise, gene flow among species was estimated from the entire sample and also for all combinations of species pairs. The models used in this study were developed for cases where equal numbers of populations occur in all races and equal numbers of races occur in all species. Unfortunately, we cannot follow this assumption because it does not reflect the perceived patterns in nature. The WEIR and COCKERHAM (1984) statistics incorporate correction terms for sample size and number of subdivisions at each level. Thus, estimates of Nm from θ may be more accurate than those based on G_{ST} which does not include correction terms.

We used mean estimates of d, n and k in formulas 3, 4 and 5 for the taxa in question. In the complete hierarchical analysis our estimates were based on $k = 3$ species, $n = 4$ races per species and $d = 100$ populations per race. Estimates of d were taken from a combination of sources, including published data on distributions (GRANT and WILKEN 1986; 1988b) and information from herbarium records. Gene flow among taxa is expressed in two ways: (1) as population-level statistics, *i.e.*, the number of migrants among populations of different races or species (Nm_{race} and Nm_{sp}, respectively) and (2) as taxon-level statistics, *i.e.*, the total number of migrants among races and species (dNm_{race} and $dnNm_{\text{sp}}$, respectively). The first set of expressions uses our estimates of k, n, and d, whereas the latter set is immune to errors in estimating the numbers of subdivisions in the hierarchy.

The second approach to estimating gene flow was the private allele method of SLATKIN (1985). Whereas WRIGHT's (1951) method takes into account differences for all alleles, SLATKIN's method uses only those alleles that are restricted to a single population. Such a "private allele" is assumed to have arisen in situ by mutation, and its frequency fluctuates by random genetic drift. If a sufficiently high frequency is reached, the allele will appear in other populations due to gene flow, at which point it loses its private status. Thus, the mean frequency of private alleles [$\bar{p}(1)$] is an indication of the average level of gene flow among populations. SLATKIN's simulation models give the formula:

$$\ln[\bar{p}(1)] \cong a\ln(Nm) + b \quad (6)$$

where a and b are constants. The model makes several assumptions, which have been examined by computer simulations (SLATKIN 1985; SLATKIN and BARTON 1989). The important assumptions are that generations are discrete, electromorphs are allelic and populations are in genetic and demographic equilibrium. In general, the relationship between $\bar{p}$ (1) and Nm is affected little by population geometry, the number of populations in existence and the number of individuals per population (SLATKIN and BARTON 1989). Relatively reliable estimates can be obtained by sampling as few as three to five populations, and the number of individuals sampled from a population affects Nm linearly so that estimates can be standardized for sample size (SLATKIN 1985).

SLATKIN's approach to estimating gene flow assumes that the private alleles sampled are uniquely derived. However, WHITKUS and CRAWFORD (1987) noted that ancestral alleles may be restricted to a single population of a species, and the frequency of such alleles may not be related to Nm in the same way as those alleles which are derived independently. To remove this effect of "historical bias," WHITKUS and CRAWFORD (1987) developed a "unique private allele" method whereby private alleles detected in related taxa are excluded from the analysis. This approach was used throughout our analysis, and the related species *I. sanctispiritus* (one population), *I. thurberi* (one population) and *I. longiflora* (two populations) were sampled to identify and exclude private alleles that were not unique to the *I. aggregata* complex.

We estimated gene flow within races and species using a nonhierarchical approach (Formula 1) and designated estimates of Nm with no subscript. This provided a means for comparing estimates from Formula 6. Within both nonhierarchical methods *I. tenuituba* was analyzed as two data sets: first as all five populations and, second, by excluding the morphologically distinct and geographically disjunct population of *I.t.* subsp. *macrosiphon*. Because of the complex statistical distributions associated with estimates of Nm, it is difficult to assess whether two Nm values are significantly different. SLATKIN and BARTON (1989) discuss the use of F statistics and private alleles and conclude that Nm estimates are probably accurate to within a factor of two. Estimates of gene flow by both types of F statistics and by private alleles were compared using Kendall's nonparametric rank correlation (SOKAL and ROHLF 1981).

To examine the relationship between geographic distance and the level of gene flow, map distances in kilometers were calculated between all pairs of populations within each subspecies of *I. aggregata*. Product-moment and Kendall's rank

TABLE 2

Estimates of gene flow (based on private alleles and nonhierarchical *F* statistics) among populations within each race of *I. aggregata*

Geographic race (subspecies) of *I. aggregata*	No. of private alleles	*Nm* from $\bar{p}(1)$	*Nm* from G_{ST}	*Nm* from θ	Distance between populations (km)			
					Minimum	Maximum	Mean	SE
aggregata	4	3.72	2.62	1.62	75	1110	615	85
formosissima (n)	0		3.12	1.38	120	545	340	30
formosissima (s)	5	4.72	1.80	1.25	130	655	475	40
attenuata	3	5.79	2.44	1.64	65	575	315	50
bridgesii	3	1.38	2.38	0.97	90	230	150	20
collina	4	6.11	2.50	1.80	35	175	100	20
candida	5	4.89	1.96	1.28	35	220	100	15
weberi	3	11.89	3.13	1.94	20	1175	605	245
Mean		5.50	2.49	1.51				

Distances between populations in each group are given in direct air kilometers, accurate to 5 km.

TABLE 3

Estimates of gene flow (*Nm*), based on nonhierarchical *F* statistics, among populations within each species of the *I. aggregata* complex

Species	No. of populations	No. of private alleles	*Nm* from $\bar{p}(1)$	*Nm* from G_{ST}	*Nm* from θ
I. aggregata	45	27	1.34	0.97	1.17
I. tenuituba[a]	5	5	0.84	0.89	0.96
	4	4	2.01	1.31	1.14
I. arizonica	6	2	6.12	1.47	1.40
Mean			2.58	1.16	1.17

[a] *I. tenuituba* was analyzed as four populations by eliminating a population of *I. tenuituba* subsp. *macrosiphon*.

correlation coefficients were calculated between *Nm* and the minimum, maximum and mean geographic distances between populations; significance of association was tested at $p = 0.05$ using the appropriate test (SOKAL and ROHLF 1981).

RESULTS

Nonhierarchical estimates of gene flow: Estimates of gene flow among populations within subspecies of *I. aggregata* were highest using private alleles and lowest using θ to calculate *Nm* (Table 2). Private alleles also gave the highest estimates of *Nm* within each of the three species when treated without subspecific classification (Table 3). Estimates of *Nm* within subspecies calculated by the private allele and θ methods were significantly rank correlated ($\tau = 0.81$; $p < 0.05$). Estimates of *Nm* from G_{ST} were similar to, but not significantly correlated with, the other two estimates. By using only unique private alleles, five otherwise private alleles were removed from the analysis of *Nm* from $\bar{p}(1)$.

In terms of the relative levels of interpopulational gene flow, SLATKIN (1981) considers values of $Nm < 1.0$ to be low because significant population differentiation can occur through drift. All but one value of *Nm* within subspecies (the θ method for *I. aggregata* subsp. bridgesii) was above 1.0; the highest estimate was for *I. aggregata* subsp. *weberi* (11.89 using private alleles; Table 2). Estimates of *Nm* within *I. aggregata* as a whole (Table 3) were lower than the mean *Nm* among populations within subspecies (Table 2). Analysis of association between estimates of gene flow and distance revealed a relationship between *Nm* and the minimum distance between populations within subspecies of *I. aggregata*. Estimates of *Nm* from θ and private alleles were both negatively correlated with minimum distance, as indicated by significant rank correlations. All other tests of association between *Nm* estimates and distance measures were not statistically significant.

Estimates of *Nm* among populations of *I. arizonica* were of similar magnitude to estimates within subspecies of *I. aggregata* (Table 3). Gene flow among the northern four populations of *I. tenuituba* was above 1.0 by both methods, and when the southern population of *I. tenuituba* subsp. *macrosiphon* was included in the analysis, *Nm* was less than 1.0 by both methods (Table 3).

Hierarchical estimates of gene flow: The hierarchical approach used here permits the comparison of gene flow rates at several taxonomic levels. In general, our estimates of Nm_{pop} were approximately 100 times higher than estimates of Nm_{race}, which were one to two times higher than estimates of Nm_{sp} (Table 4). Estimates of Nm_{pop} were of similar magnitude to nonhierarchical estimates in Tables 2 and 3. On average, hierarchical values calculated from G_{ST} were higher than those from θ, but estimates by the two methods were significantly rank correlated.

Estimates of total gene flow among races (dNM_{race}) were of similar magnitude to Nm_{pop}. However, dNm_{race} for *I. aggregata* was at least three times that for *I. tenuituba*. Estimates of total gene flow among species ($dnNm_{sp}$) were two to three times higher than dNm_{race}. Some unexpected patterns were seen in the analysis of gene flow among species pairs. The estimate of $dnNm_{sp}$ between *I. arizonica* and sympatric populations

TABLE 4

Estimates of gene flow based on hierarchical *F* statistics

Level	Taxon (taxa)[a]	From G_{ST}	From θ
Nm_{pop}	*I. aggregata*	1.50	1.32
	I. tenuituba	1.14	1.06
	All taxa	1.33	1.12
Nm_{race}	*I. aggregata*	0.024	0.018
		(2.40)	(1.80)
	I. tenuituba	0.008	0.005
		(0.80)	(0.50)
	All taxa	0.017	0.010
		(1.68)	(1.00)
Nm_{sp}	*aggregata*$_7$ × *tenuituba*	0.042	0.031
		(16.80)	(12.40)
	aggregata$_{45}$ × *tenuituba*	0.048	0.037
		(19.20)	(14.80)
	aggregata$_7$ × *arizonica*	0.054	0.043
		(21.60)	(17.20)
	aggregata$_{45}$ × *arizonica*	0.031	0.017
		(12.40)	(6.60)
	tenuituba × *arizonica*	0.019	0.009
		(7.60)	(3.60)
	All three species	0.010	0.006
		(9.20)	(4.80)

Values given are for the number of migrants between populations at each level; values in parentheses are estimates for total numbers of migrants among taxa (dNm_{race} and $dnNm_{sp}$).

[a] All five populations of *I. tenuituba* were used in the hierarchical analysis.

of *I. aggregata* (*aggregata*$_7$) were the highest estimates of interspecific gene flow. This is somewhat surprising given that there is no morphological evidence of hybridization between these species in nature, whereas *I. aggregata* and *I. tenuituba* are known to hybridize, yet had similar estimates of $dnNm_{sp}$ than did *I. arizonica* and *I. aggregata*.

DISCUSSION

Gene flow within taxa: Estimates of interpopulational gene flow within *I. aggregata* were generally high; the data suggest that reproductive continuity is sufficient to prevent allelic fixation in different populations. Estimates of *Nm* among populations within races of *I. aggregata* were between about 1 and 12 for distances ranging from 22 to over 1000 km. In Colorado, the main visitors to *I. aggregata* are the broad-tailed hummingbird (*Selasphorus platycerus*) and the rufous hummingbird (*S. rufus*). These birds can move extremely long distances during migration; one of the longest recorded movements is for a female rufous hummingbird banded in Montana and recaptured 1202 km south in Colorado (CALDER and JONES 1989). Although this was probably not a nonstop flight (W. CALDER, personal communication) it illustrates the potential for long-distance gene flow because *I. aggregata* was one of the major food sources at the time of this flight and Ipomopsis pollen can remain viable for several days (N. WASER, unpublished data). Patterns of hummingbird migration also suggest that gene flow may, in general, follow a stepping-stone model of population structure rather than an island one.

Nonhierarchical estimates of *Nm* among all populations of *I. tenuituba* (including subsp. *macrosiphon*) were less than 1.0 using both methods. This pattern was also reflected in our hierarchical analysis whereby dNm_{race} was considerably higher for *I. aggregata* than for *I. tenuituba*. Evidence from allozymic and chloroplast DNA variation (WOLF 1990; WOLF, SOLTIS and SOLTIS 1991) suggests that subsp. *macrosiphon* was derived independently from a related species, *I. thurberi*, whereas other populations of *I. tenuituba* are more closely related to *I. aggregata*. Thus, all five populations of *I. tenuituba* apparently represent an artificial assemblage of morphologically similar but genealogically distinct populations, and high levels of gene flow among them should not necessarily be expected.

Gene flow among geographic races: PORTER (1990) provides a qualitative interpretation for intertaxon estimates of gene flow (Nm_{GT}). It should be noted that the "*N*" in this context is equivalent to the effective population size of the taxon (*i.e.*, *dN* for races and *dnN* for species). PORTER (1990) suggests that values above unity are the result of significant genetic similarity among taxa due to gene flow. When $\sim 1.0 > Nm_{GT} > \sim 0.5$, gene flow is weak but possibly effective for the transfer of favorable genes between taxa, and values below ~0.5 suggest that groups are almost or fully isolated. The assumption here is that the effect of dNm_{race} and $dnNm_{sp}$ on the differentiation of races and species, respectively, is analogous to the effect of *Nm* on the differentiation of populations within a taxon. At the racial level of our analysis, only gene flow among races of *I. tenuituba* could be considered low, but still sufficient to allow transfer of favorable alleles, according to PORTER's (1990) criteria.

Patterns of morphological divergence in *I. aggregata* suggest that geographic speciation is incipient. Our estimates of gene flow among races appear high enough to prevent a strong effect of genetic drift at neutral loci. However, the designation of geographic races as subspecies is somewhat arbitrary because in most cases the morphological forms intergrade across the boundaries. In this study we avoided morphologically intermediate populations, and thus, populations at the range peripheries are probably exchanging genes at levels higher than indicated by our estimates of dNm_{race}. Nevertheless, it appears that morphological differences are maintained despite apparent gene flow between races.

Although we treat *I. aggregata* subsp. *candida* as a geographic race, GRANT and WILKEN (1986) describe it as a more divergent taxon: a semispecies. GRANT

and WILKEN (1988a) hypothesize that subsp. *candida* may have once been a separate species that has recently come into secondary contact with *I. aggregata* subsp. *collina* through habitat disturbance and breakdown of ecological isolation. This is supported by ELAM and LINHART (1988) who failed to detect ethological isolation between these taxa as a result of pollinator preferences. An alternative hypothesis to secondary integradation is that subspp. *candida* and *collina* are undergoing primary divergence (WILKEN and ALLARD 1986). Results from molecular studies suggest these two subspecies are sister taxa (WOLF 1990). Thus, unless a "primitive" *candida* chloroplast genome has been lost in all populations sampled, molecular data support primary divergence, at least with respect to the rest of *I. aggregata*. However, it is impossible to rule out some form of secondary contact with the available data.

Gene flow among species: On average we estimated approximately five to ten individuals migrating between species of the *I. aggregata* complex. One of the factors which we expect to limit interspecific gene flow is differences in floral morphology as a result of adaptations to pollinators. PAIGE and WHITHAM (1985) found that hawkmoths prefer to visit white flowers (of *I. tenuituba*) and hummingbirds prefer red flowers (*I. aggregata*). In some areas these differences break down, and *I. aggregata* and *I. tenuituba* hybridize, as reflected in our estimates of $dnNm_{sp}$. However, *I. arizonica* is not known to hybridize with other taxa, yet our estimates of interspecific gene flow involving this species were high, especially between *I. arizonica* and sympatric populations of *I. aggregata*. The primary morphological difference between *I. arizonica* and *I. aggregata* is the position of anthers, which are included in *I. arizonica* and exserted in *I. aggregata*. In their area of sympatry, both species have red flowers and are visited by hummingbirds which occasionally visit both species during a single foraging bout at the contact zone (P. WOLF personal observation). GRANT and WILKEN (1986) propose that hummingbirds receive and deliver *arizonica* pollen on the bill, whereas *aggregata* pollen is transferred on the face feathers, resulting in positive assortative mating when both species are visited. Experiments are underway to examine the effect of different floral morphologies on bird-mediated pollen transfer between taxa, but preliminary data indicate that hummingbirds can transfer pollen between *I. aggregata* and *I. arizonica* (P. WOLF and D. CAMPBELL, unpublished data). This suggests that factors other than anther position are responsible for an apparent lack of natural hybrids. At least two hypotheses can explain a high estimate of gene flow between *I. aggregata* and *I. arizonica*. First, *I. arizonica* has diverged from *I. aggregata* very recently and our estimates include levels of gene flow prior to complete isolation. However, phylogenetic analysis of chloroplast DNA suggests that *I. arizonica* has not been derived from within *I. aggregata* but is a sister taxon to it (WOLF 1990), suggesting a more ancient divergence. Alternatively, introgression may have occurred but has not been detected morphologically. D. WILKEN (personal communication) has found that several floral traits in Ipomopsis are under relatively simple genetic control. In addition, some floral characters can experience strong phenotypic selection during pollination (CAMPBELL 1989). Thus, differences in floral morphology could evolve rapidly, and, if selection is strong, differences could be maintained even if gene flow is high. Thus, morphological hybrids would not be detected a few generations following hybridization, but "alien" neutral genes would remain. Such patterns of pollen flow might not be revealed by chloroplast DNA which is maternally inherited and not transferred by pollen (WOLF 1990).

Assumptions of models: The main underlying assumption of the models used in this study is that allele frequencies are directly related to the level of gene flow. Thus, the models assume that an equilibrium has been reached between genetic drift and migration. Detailed assumptions of both F statistics and private allele models are discussed by SLATKIN and BARTON (1989), and assumptions for using f statistics for estimates of gene flow among taxa are considered by PORTER (1990) and SLATKIN and VOELM (1991). The effects of selection and mutation on the models are likely to be minimal. Diversifying selection will lead to an overestimate of Nm, and balancing selection will result in an underestimate of Nm. In practice, many loci are used (23 in the present study), and average effects are lower than single-locus and single-allele effects. Mutation is usually at such a low background level that it will have little effect on estimates of Nm unless Nm itself is small (SLATKIN 1985).

Gene flow may be overestimated by non-hierarchical methods if the number of populations in existence is less than about 10 (NEI, CHAKRAVARTI and TATENO 1977). In our study this is the case only for *I. tenuituba* subsp. *macrosiphon*, but we did not estimate Nm among populations of this taxon except within the hierarchical analysis. The significant negative correlation of Nm with minimum geographic distance suggests that gene flow is highest among geographically adjacent populations and that overall estimates of Nm are biased by closely sampled populations. Collecting for a study of gene flow should therefore attempt to sample widely and evenly, and avoid sampling two or more populations in relatively close proximity.

Overestimation of Nm can result from the occurrence of hidden alleles, *i.e.*, those encoding apparently identical electromorphs. To maximize resolution of interpopulational allozymic comparisons we used

more than one buffer system for some enzymes (WOLF, SOLTIS and SOLTIS 1991). We observed a total of 98 allozymes in this study, but only knowledge of the encoding DNA sequences can determine the total number of truly different alleles.

If populations have diverged recently, or colonizations and extinctions have been frequent, genetic drift and migration may not have reached equilibrium, resulting in overestimates of *Nm* (*e.g.*, LARSON, WAKE and YANEV 1984). Currently, populations of the *I. aggregata* complex inhabit high-elevation "islands" throughout western North America. These habitats were more continuous during the Wisconsin glaciation, *ca.* 12,000 years b.p. (SPAULDING, LEOPOLD and VAN DEVENDER 1983), and the distribution of Ipomopsis was probably more continuous at that time (GRANT and WILKEN 1988b). Furthermore, populations in the northern part of the range have probably invaded the area in the last 10,000 years, following glacial retreat, and thus, present-day populations may have radiated from a few common source populations. These changes in distribution may reflect demographic instability. The effect of this type of history on estimates of gene flow was examined by SLATKIN (1985) who showed that the relationship between private-allele frequency and *Nm* holds if the number of generations since divergence is greater than the effective population size. SLATKIN and BARTON (1989) have extended this assumption to estimates using *F* statistics. The average generation time in *I. aggregata* is approximately four years (WASER and PRICE 1989), and the effective population size is probably less than 100 (CAMPBELL 1989). Thus, glacial history has probably had little effect on our estimates of *Nm*. The removal of private ancestral alleles from our analyses of $\bar{p}(1)$ also contributed to reducing the risks of historical bias. Models for estimating *Nm* among taxa using *F* statistics have been examined by PORTER (1990) and SLATKIN and VOELM (1991). However, simulation studies are needed to test the robustness and sensitivity of these models to historical events.

LEVIN (1988) discussed the effects of various factors, such as sib-structured migration and variance in the level of gene flow, that may lead to inaccurate estimates. Clearly, gene flow is an extremely complex phenomenon that can occur in many ways, and can fluctuate over time. Thus the values of *Nm* that we present should be taken as general estimates of "effective gene flow" as if populations were of similar sizes, and migrants were unrelated and equally likely to move to any other population. Only such a general statistic can be used for comparative purposes and only by comparing such estimates to measures of realized gene flow can the assumptions of the models be tested rigorously.

Comparisons of methods: SLATKIN and BARTON (1989) showed that *F* statistics and private alleles provide roughly equal estimates of *Nm* under a wide variety of conditions, but *F* statistics are preferred because they use all alleles and reduce the risks of misinterpreting a few. Differences between the estimators of F_{ST} are minimal, except when *Nm* is high (approximately five times the highest estimate in our study), whereby *Nm* is overestimated using θ, and underestimated using G_{ST}. PORTER (1990) found that estimates of *Nm* among sympatric sibling species using G_{ST} were too high. We also found estimates of *Nm* to be higher from G_{ST} than θ, but not to the same degree as PORTER (1990). Correction terms for sample size may be responsible for the same patterns seen in both studies, suggesting that θ may be a more reliable estimator for hierarchical analyses of gene flow.

Comparisons between direct and indirect estimates of *Nm* may be used to examine the validity of the indirect models, although such comparisons should not ignore temporal fluctuations that point estimates can miss. Pollen flow among local populations of *I. aggregata* subsp. *aggregata* in Colorado fluctuates from year to year and varies among individuals (WASER and PRICE 1983; CAMPBELL and WASER 1989). The above studies used measures of pollen dispersal, which CAMPBELL (1991) has shown to occur over considerably shorter distances than realized gene flow, as estimated from paternity exclusion analysis. CAMPBELL (1991) found that at least 16% of seeds were sired by plants outside a 10 × 10-m square surrounding each plant and in one of the populations at least 4% were pollinated by plants outside a 30 × 30-m plot (CAMPBELL 1991). These are minimum estimates because they are not corrected for cryptic gene flow (DEVLIN and ELLSTRAND 1990). Although CAMPBELL's study does not provide values of *Nm*, it does demonstrate that long-distance pollen delivery in this species is high. By increasing the scale of analysis to perhaps a few kilometers and using paternity analyses over several years, it should be possible to make estimates of *Nm* directly from realized gene flow. These values could then be compared to indirect estimates to test the indirect models at least at a local level.

Conclusions: The amount of gene flow required to prevent speciation will depend on the strength of other evolutionary forces. If our estimates of *Nm* approach the actual levels of gene flow in the *I. aggregata* complex, it appears that speciation has occurred even though gene flow is strong enough to offset the effect of genetic drift. This is possible if selection can counteract the effect of gene flow (ENDLER 1977). Gene flow, like other forces, is higher in some parts of the genome and lower in others. For example, favorable genes can still be exchanged successfully even when barriers to gene flow are strong (BARTON and BENGTSSON 1986). The opposite effect,

whereby unfavorable genes are not exchanged under high gene flow, is also possible. Such genes could be at loci that confer local adaptations and at any linked loci. The significance of this is that *Nm*, even if estimated accurately, may still fail to account for variation among different parts of the genome.

I. aggregata is probably still undergoing geographic speciation, whereas *I. arizonica* is so distinct morphologically that it has been described as the "best biological species in the group" (GRANT and WILKEN 1986). The reduced fertility of F_1 hybrids between *I. aggregata* and *I. arizonica* suggests that isolating mechanisms are evolving. *I. aggregata* and *I. arizonica* may have diverged initially under an ecologically and geographically parapatric (or allopatric) speciation pattern. Gene flow between the two taxa may have ceased temporarily, long enough for ecological, and perhaps morphological, divergence to result under selection, but not long enough for genetic drift to produce a detectable effect at neutral loci. The alternative, that allozymic similarities between the two taxa are the result of similar selection pressures, seems unlikely given the differences between the habitats.

In summary, indirect estimates of *Nm* suggest that gene flow among populations within species is generally high. Genetic exchange between geographic races of *I. aggregata* is also high. High gene flow estimates between *I. aggregata* and *I. tenuituba* were expected because these species are known to hybridize in nature. However, contrary to morphological evidence we also found relatively high rates of interspecific gene flow involving *I. arizonica*. In Ipomopsis, it appears that speciation has occurred without the complete cessation of gene flow between diverging taxa.

Thanks to DIETER WILKEN for help in collecting material. Specimens used for locality information were borrowed from RSA/POM, RM and MNA. Thanks to ROBIN BUSH, DIANE CAMPBELL, DICK HUDSON, STEVE NOVAK, ADAM PORTER, LOREN RIESEBERG, DOUG SOLTIS, JOHN THOMPSON, NICK WASER and RICK WILLIAMS for comments on various versions of the manuscript. This research was funded by grants from Sigma Xi, The Hardman Foundation, a Washington State University Summer Research Assistantship and National Science Foundation Dissertion Improvement Grant BSR-8800978.

LITERATURE CITED

BARTON, N. H., and B. O. BENGTSSON, 1986 The barrier to genetic exchange between hybridizing populations. Hereditary **57:** 357–376.

BUTLIN, R. K., and AND G. M. HEWITT, 1985 A hybrid zone between *Chorthippus parallelus parallelus* and *Chorthippus parallelus erythropus* (Orthoptera: Acrididae): morphological and electrophoretic characters. Biol. J. Linn. Soc. **26:** 269–285.

CALDER, W., III, and AND E. G. JONES, 1989 Implications of recapture data for migration of the rufous hummingbird (*Selasphorus rufus*) in the Rocky Mountains. Auk **106:** 488–489.

CAMPBELL, D. R., 1989 Measurements of selection in a hermaphroditic plant: variation in male and female pollination success. Evolution **43:** 318–334.

CAMPBELL, D. R, 1991 Comparing pollen dispersal and gene flow in a natural population. Evolution (in press).

CAMPBELL, D. R., and N. M. WASER, 1989 Variation in pollen flow within and among populations of *Ipomopsis aggregata*. Evolution **43:** 1444–1455.

DEVLIN, B., and N. C. ELLSTRAND, 1990 The development and application of a refined method for estimating gene flow from angiosperm paternity analysis. Evolution **44:** 248–259.

EHRLICH, P., and P. RAVEN, 1969 Differentiation of populations. Science **165:** 1228–1232.

ELAM, D. R., and Y. B. LINHART, 1988 Pollination and seed production in *Ipomopsis aggregata*: differences among and within flower color morphs. Am. J. Bot. **75:** 1262–1274.

ENDLER, J. A., 1977 *Geographic Variation, Speciation, and Clines.* Princeton University Press, Princeton, N.J.

FELSENSTEIN, J., 1981 Skepticism towards Santa Rosalia, or why are there so few kinds of animals? Evolution **35:** 124–138.

GRANT, V., 1981. *Plant Speciation.* Columbia University Press, New York, N.Y.

GRANT, V., and D. H. WILKEN, 1986 Taxonomy of the *Ipomopsis aggregata* group (Polemoniaceae). Bot. Gaz. **147:** 359–371.

GRANT, V., and D. H. WILKEN, 1988a Natural hybridization between *Ipomopsis aggregata* and *I. tenuituba* (Polemoniaceae). Bot. Gaz. **149:** 213–221.

GRANT, V., and D. H. WILKEN, 1988b Racial variation in Ipomopsis tenuituba (Polemoniaceae). Bot. Gaz. **149:** 443–449.

HAMRICK, J. L., 1987 Gene flow and distribution of genetic variation in plant populations, pp. 53–67 in *Differentiation Patterns in Higher Plants*, edited by K. M. Urbanksa Academic Press, Orlando, FL.

HARRISON, R. G. 1979. Speciation in North American Field crickets: evidence from electrophoretic comparisons. Evolution 33:1009–1023.

HEWITT, G. M. 1989. The subdivision of species by hybrid zones, pp. 85–110. In D. OTTE and J. A. ENDLER (eds.) Speciation and its Consequences. Sinauer Associates, Inc. Publishers, Sunderland, MA.

LARSON, A., D. B. WAKE and K. P. YANEV. 1984 Measuring gene flow among populations having high levels of genetic fragmentation. Genetics **106:** 293–308.

LEVIN, D. A., 1975 Interspecific hybridization, heterozygosity and gene exchange in *Phlox*. Evolution **29:** 37–51.

LEVIN, D. A., 1988 Consequences of stochastic elements in plant migration. Am. Nat. **132:** 643–651.

MALLET, J., N. BARTON, M. G. LAMAS, C. J. SANTISTEBAN, M. M. MEUDAS and H. EELEY, 1990 Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in Heliconius hybrid zones. Genetics **124:** 921–936.

MAYR, E., 1963 *Animal Species and Evolution.* Harvard University Press, Cambridge, Mass.

MAYR, E., 1970 *Populations, Species, and Evolution.* Harvard University Press, Cambridge, Mass.

NEI, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

NEI, M., A. CHAKRAVARTI and Y. TATENO, 1977 Mean and variance of F_{ST} in a finite number of incompletely isolated populations. Theor. Popul. Biol. **13:** 291–306.

PAIGE, K. N., and T. G. WHITHAM, 1985 Individual and population shifts in flower color by scarlet gilia: a mechanism for pollinator tracking. Science **227:** 315–317.

PATERSON, H. E. H. 1985. The recognition concept of species, in *Species and Speciation* (Transvaal Museum Monograph No. 4) edited by E. S. VRBA. Pretoria.

PORTER, A., 1990 Testing nominal species boundaries using gene flow statistics: the taxonomy of two hybridizing admiral butterflies (*Limenitis:* Nymphalida). Syst. Zool. **39:** 131–147.

SLATKIN, M., 1981 Estimating levels of gene flow in natural populations. Genetics **99:** 323–335.

SLATKIN, M., 1985 Rare alleles as indicators of gene flow. Evolution **39:** 53–65.

SLATKIN, M., 1987 Gene flow and the geographic structure of natural populations. Science **236:** 787–792.

SLATKIN, M., and N. H. BARTON, 1889 A comparison of three methods for estimating average levels of gene flow. Evolution **43:** 1349–1368.

SLATKIN, M., and L. VOELM, 1991 F_{ST} in a hierarchical island model. Genetics **127:** 627–629.

SOKAL, R. R., and R. J. ROHLF, 1981 *Biometry*, Ed. 2. Freeman, San Francisco.

SPAULDING, W. G., E. B. LEOPOLD and T. R. VAN DEVENDER, 1983 Late Wisconsin paleoecology of the American southwest, pp. 259–293 in *Late-Quaternary Environments of the United States*, Vol 1, edited by H. E. WRIGHT. University of Minnesota Press, Minneapolis.

STEBBINS, G. L., 1950 *Variation and Evolution in Higher Plants.* Columbia University Press, New York.

STREAMS, F. A., and D. PIMENTAL, 1961 Effects of immigration on the evolution of populations. Am. Nat. **95:** 201–210.

SWOFFORD, D. L., and R. K. SELANDER. 1981 BIOSYS-1. A FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. J. Hered. **72:** 281–283.

TEMPLETON, A. R., 1989 The meaning of species and speciation, pp. 3–27 in *Speciation and Its Consequences*, edited by D. OTTE and J. A. ENDLER. Sinauer Associates, Sunderland, Mass.

THODAY, J. M., and T. B. BOAM, 1959 Effects of disruptive selection. II. Polymorphism and divergence without isolation. Heredity **13:** 205–218.

WASER, N. M., and M. V. PRICE, 1983 Optimal and actual outcrossing, and the nature of plantpollinator interaction, pp. 341–359 in *Handbook of Experimental Pollination Biology*, edited by C. E. JONES and R. J. LITTLE. Van Nostrand Reinhold, New York.

WASER, N. M., and M. V. PRICE, 1989 Optimal outcrossing in *Ipomopsis aggregata*: seed set and offspring fitness. Evolution **43:** 1097–1109.

WEIR, B. S., and C. C. COCKERHAM, 1984 Estimating *F*-statistics for the analysis of population structure. Evolution **38:** 1358–1370.

WHITKUS, R., and D. J. CRAWFORD, 1987 Estimates of gene flow–the importance of historical distributions. Am. J. Bot. **74:** 671–672.

WILKEN, D. H., and S. T. ALLARD, 1986 Intergradation among populations of the *Ipomopsis aggregata* complex in the Colorado Front Range. Syst. Bot. **11:** 11–13.

WOLF, P. G., 1990 Molecular studies of incipient speciation in the *Ipomopsis aggregata* complex (Polemoniaceae). Ph.D. dissertation, Washington State University, Pullman.

WOLF, P. G., P. S. SOLTIS and D. E. SOLTIS, 1991 Genetic relationships and patterns of allozymic divergence in the *Ipomopsis aggregata* complex (Polemoniaceae). Am. J. Bot. **78:** 515–526.

WRIGHT, S. 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.

WRIGHT, S. 1951 The genetical structure of populations. Ann. Eugenics **15:** 323–354.

WRIGHT, S., 1965 The interpretation of population structure by *F*-statistics with special regard to systems of mating. Evolution **19:** 355–420.

WRIGHT, S., 1978 *Evolution and the Genetics of Populations, Vol. 4. Variability Within and Among Natural Populations.* University of Chicago Press, Chicago.

Communicating editor: R. R. HUDSON